



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

MODELO SOBREVIDA PARA LA DESERCIÓN ESTUDIANTEL DE LOS PROGRAMAS DEL NIVEL TÉCNICO EN UNA IES DE FORMACIÓN PARA EL TRABAJO EN LA CIUDAD DE BOGOTÁ

SURVIVAL MODEL FOR STUDENT DROPOUT FROM THE TECHNICAL LEVEL PROGRAMS IN A WORK TRAINING HEI IN THE CITY OF BOGOTÁ

Wilmer Rene Matallana Alfonso

wrmatalanaa@libertadores.edu.co

John González Veloza

jjgonzalezv02@libertadores.edu.co

Lida Rubiela Fonseca Gómez

lrfonsecag@libertadores.edu.co

Fundación Universitaria Los Libertadores

RESUMEN

El presente trabajo, tiene como finalidad abordar la deserción estudiantil en una IES de Formación para el Trabajo en la ciudad de Bogotá, en general, se pretende explicar, interpretar y documentar, cuáles fueron las principales causas que llevaron a la deserción, a este grupo de estudiantes de la oferta académica 2019 de IES en mención, esto con el fin de tomar decisiones desde lo administrativo mitigando esta problemática, en pro del beneficio de los estudiantes, colaboradores y de la misma institución.

Para lo cual se determinó un modelo sobrevivencia que explique esta problemática, en este modelo, se realiza un análisis para cada individuo durante un determinado período y en los que se establece el evento que transcurre entre el evento inicial y el evento final o hasta que acaba el seguimiento si no ocurre el evento final. Este análisis de supervivencia se realizó bajo las metodologías de Kaplan Meier y el modelo de riesgo proporcional de Cox.

El seguimiento se realizó a un total de 510 estudiantes en un periodo comprendido entre el 26 de marzo de 2019 y 26 de marzo de 2020, de los cuales 348 eran mujeres y 162



hombres, este grupo de estudiantes se encuentran en los estratos 1 y 2, y de los cuales 96 desertaron equivalente al 19%.

De lo cual, se pudo establecer que dentro de las principales causas por la que este grupo de estudiantes abandonaron la formación, están relacionadas si los estudiantes se encontraban trabajando o no, teniendo en cuenta que, los estudiantes sin trabajo tienen una probabilidad del 79%, mientras que los que tienen trabajo del 82% de abandonar los estudios.

Respecto a la Alternativa de Etapa Productiva seleccionada y teniendo en cuenta las garantías económicas que cada una de estas puede ofrecer, se evidencio que, los aprendices que seleccionaron Vínculo laboral, desertaron hasta el quinto mes antes de iniciar la etapa productiva y se mantuvo constante hasta la finalización de la formación con un porcentaje de supervivencia del 40%, los estudiantes que no seleccionaron ninguna alternativa desertaron desde el primer hasta el cuarto mes de formación.

Palabras clave: Deserción, Análisis de Supervivencia, Estimación Kaplan-Meier, modelo Regresión de Cox.

ABSTRACT

The purpose of this work is about students' desertion at the IES Job Training in Bogotá. It is intended to explain, interpret and document, what were the main causes that led to desertion, the students' group in the academic offer 2019 at the IES, in order to take administrative decisions mitigating this problem, to the students' benefit, collaborators and the institution itself.

was established a survival model to explain this problem, in this model is carried out an analysis for each people during a certain period and in which event elapses between the initial event and the final event or until it follows up to end or it does not occur the final event. This survival analysis was performed using the Kaplan Meier methodologies and the Cox proportional hazard model.

A follow-up was carried out in a total of 510 students in a period between March 26 in, 2019 and March 26 in 2020, there were 348 women and 162 men, this group of students are stratified in strata one and strata two, and of which 96 abandoned the process and it is equivalent to 19%.

It could be established that among the main causes that explain the reason of students 'group dropped out of training are related to the students were working or not, taking into account if the students does not have a job the probability to remain in the process it was 79%, while the people has a job they abandon the process in 82%.

Keywords: Attrition, Survival Analysis, Kaplan-Meier Estimate, Cox Regression model



INTRODUCCIÓN

Este estudio surge como respuesta a la necesidad de desarrollar un modelo sobrevida para conocer el comportamiento del fenómeno de la deserción en los programas de nivel Técnico de una entidad de formación para el trabajo en la ciudad de Bogotá, la información para la realización del análisis, es proporcionada por la oficina de Bienestar, lo cual, permite que se efectúe a través de datos confiables que deriven en mejores estrategias para su prevención, dado que los modelos y sistemas de información actuales no permiten una adecuada lectura institucional sobre el fenómeno y no es posible adaptarlos al contexto de la formación y la situación de los estudiantes. (Arteaga González, Torres Ávila, & Lopez Cardona, 2016, pág. 65).

En las Entidades de Formación, independiente la modalidad o el nivel, la deserción escolar es una problemática que deben afrontar en cada uno de sus programas, y en cada una de sus ofertas académicas, por tanto, este inconveniente es constante, que no genera dificultad solamente en la estabilidad de la entidad, si no también incertidumbre entre sus docentes y demás colaboradores, de acuerdo con (Pineda, Pedraza, Baquero, Dussán & Ramírez, 2010). (Gartner Isaza, 2020). En la perspectiva nacional se considera que la deserción compromete la estabilidad de los ingresos de las instituciones educativas y la calidad de sus procesos administrativos y académicos, por cuanto constituye un costo económico y social que es asumido por las familias, las universidades y el Estado.

La Entidad de Formación para el Trabajo del nivel técnico, ubicada en la localidad de Usme en la Ciudad de Bogotá, brinda apoyo a sus estudiantes desde el componente de bienestar al aprendiz, como “estrategia institucional para contribuir en la permanencia y el desempeño exitoso de los aprendices de la entidad en su proceso formativo con enfoque territorial y diferencial, con el que se busca ofrecer a los aprendices oportunidades para el fortalecimiento de su desarrollo humano integral mediante: la implementación de sus programas, proyectos y actividades para el acompañamiento y reconocimiento de sus capacidades y talentos; que permitan enfrentar con éxito las exigencias del mundo de la vida en el marco de la formación y su incorporación al mundo del trabajo” (Servicio Nacional de Aprendizaje, 2020)

De acuerdo con lo anterior, se hace necesario realizar estudios y ejercicios con el fin de minimizar la incidencia del ausentismo escolar, para que desde lo institucional se puedan tomar las medidas necesarias, contrarrestando esta problemática.

Es así, que después de verificar diferentes autores e investigaciones, en el estudio y análisis de la problemática de la deserción, la mejor opción son los análisis de sobrevida, la cual, permite establecer la ocurrencia de un evento en un determinado espacio de tiempo permitiendo detectar los posibles factores que conllevan a la decisión del ausentismo escolar.



De acuerdo con Cárdenas Leuro, (2013) “La sobrevida o tiempo de supervivencia se entiende como el período de tiempo desde el inicio de la observación hasta que un evento ocurre. Las funciones de sobrevida modelan el tiempo hasta la ocurrencia de un evento, entre otros, la muerte. En general el tiempo es una variable de tipo continuo en cualquier unidad como mes, año, día, segundos, etc., medida desde el inicio de la observación del individuo en estudio hasta que el evento de interés ocurre.”

Por otra parte Leidy Alejandra Ocampo Mesa, (2019) manifiesta que “el análisis de supervivencia constituye una serie de procedimientos y técnicas estadísticas para analizar datos en los cuales la variable de interés representa el tiempo de duración de un evento (El termino supervivencia se debe a que la técnica se utilizó inicialmente en aplicaciones donde el evento de interés es la muerte) Barrera Rebellon, (2008),” donde la ocurrencia del evento para efectos de esta investigación es considerada como la deserción de los estudiantes del nivel técnico de una entidad de formación para el trabajo en la ciudad de Bogotá, en un periodo de tiempo entre 26 de marzo de 2019, momento en el cual inicial la formación hasta el 26 de marzo de 2020, tiempo de finalización del proceso académico, “de esta forma el concepto de datos censurados propios de la teoría de sobrevida es de gran importancia dentro de la investigación ya que se presentan cuando hasta la última observación que se le hace a los estudiantes no se presenta en evento objeto de estudio, es decir que son todos aquellos estudiantes que se gradúan del programa técnico, de esta forma el uso la función de sobrevivencia proporcionará la probabilidad de que al estudiante no le ocurra el evento de observación (deserción).”.

Para estimar la función de sobrevida es necesario implementar el concepto de Kaplan y Meier, debido al tipo de censura mencionado anteriormente.

El estimador de Kaplan y Meier (1958), de la función de supervivencia, es el estimador más común entre los paquetes estadísticos. El método de construcción del estimador es un método no paramétrico ya que no asume ninguna estructura para la función de distribución del tiempo de vida, permitiendo comparar funciones de sobrevida obteniendo factores influyentes en la decisión de abandonar los estudios. Barrera Rebellon, (2008)

Por último se elaborará el modelo semiparamétrico de regresión Cox, que de acuerdo a Velasco Álvarez, (2015/16) el modelo parte de una hipótesis fundamental, la de que los riesgos son proporcionales y es utilizado para detectar relaciones existentes entre el riesgo que se produce en un determinado individuo en el estudio y algunas variables independientes y/o explicativas; por lo que nos permite evaluar dentro de un conjunto de variables cuáles tienen relación, influencia...sobre la función de riesgo y por ello también en la función de supervivencia, ya que ambas funciones están conectadas.

Por tanto, este modelo nos permite detectar el riesgo de abandono de estudios asociadas a las variables de mayor relevancia, debido a que se buscará la relación entre los riesgos de deserción de los individuos expuestos a factores de riesgo diferentes.



REFERENTES TEORICOS

En Colombia la educación se define como un proceso de formación permanente, personal cultural y social que se fundamenta en una concepción integral de la persona humana, de su dignidad, de sus derechos y de sus deberes, el sistema educativo colombiano lo conforman: la educación inicial, la educación preescolar, la educación básica (primaria cinco grados y secundaria cuatro grados), la educación media (dos grados y culmina con el título de bachiller), y la educación superior "Sistema educativo colombiano". Autor: Ministerio. De: Argentina. Para: Concepto. de. Disponible en: <https://concepto.de/educacion-4/>. Consultado: (26 de agosto de 2020).

Ahora bien, después de contextualizado el proceso de la educación, entramos a definir que es la deserción escolar, la cual se trata generalmente de un conjunto de causas sociales, culturales y económicas que confluyen para que los estudiantes abandonen la escuela y se dediquen a otra cosa. Deserción Escolar". Autor: María Estela Raffino. De: Argentina. Para: Concepto. (2020).

La deserción no se da simplemente porque las personas no quieren estudiar, sino que es un fenómeno complejo que delata otras razones más profundas en la sociedad, la cual puede entenderse como una situación a la que se enfrenta un estudiante cuando aspira concluir su proyecto educativo y no lo logra. Se considera desertor a aquel individuo que siendo estudiante de una institución de educación superior no presenta actividad escolar durante dos semestres académicos consecutivos, lo cual equivale a un año de inactividad académica Tinto, 1982; Tinto & Wallace, (1986); (Men, 2009).

La deserción definida desde el individuo implica reconocer que los significados que un estudiante asigna a su comportamiento de abandono pueden diferir sustancialmente de los que un observador atribuye a ese mismo comportamiento. Desertar significa entonces el fracaso individual en completar un determinado curso de acción para alcanzar una meta deseada. Universidad de los Andes, (2014), citado por Gartner Isaza, Lorena y Dussán Lubert, Carmen y Montoya, Diana Marcela (2016). Caracterización de la Deserción Estudiantil en la Universidad de Caldas el Período (2009-2013)

Teniendo en cuenta lo anterior, es necesario aclarar que la deserción escolar se puede presentar en varios tipos:

- ☐ Deserción precoz. Cuando el estudiante ha sido aceptado en un programa escolar y no acude nunca al centro educativo, ni completa sus clases.
- ☐ Deserción temprana. Cuando el estudiante abandona el programa de estudios durante los cuatro primeros semestres.
- ☐ Deserción tardía. Cuando el estudiante abandona el programa de estudios del quinto semestre en adelante.



- Deserción total. Cuando el estudiante abandona por completo un plan educativo y no regresa a él nunca.
- Deserción parcial. Cuando el estudiante se da una baja temporal de una duración determinada y luego retoma sus estudios. Deserción Escolar". Autor: María Estela Raffino. De: Argentina. Para: Concepto (2020).

Dentro de las técnicas más utilizadas y que brindan gran confiabilidad a la hora de obtener resultados principalmente para la toma de decisiones, encontramos el Análisis de Supervivencia, la cual es una técnica inferencial cuyo principal objetivo es examinar y modelar el tiempo que se toma para que ocurra un determinado suceso, la función de supervivencia se define como la probabilidad de que una persona sobreviva (no le ocurra el evento de interés) al menos hasta el tiempo t . Sea T una variable aleatoria continua positiva (o no negativa) con función de densidad de probabilidad $f(t)$ y función de distribución acumulada $F(t) = P(T \leq t)$, la función de supervivencia $S(t)$ se define como $S(t) = 1 - F(t) = P(T > t)$, que representa la probabilidad de que se produzca el evento en el tiempo mayor que t .

Este análisis de supervivencia se realizará bajo las metodologías de Kaplan Meier, y el modelo de riesgo proporcional de Cox, que generalmente, dada las innumerables aplicaciones de estas técnicas en el ámbito biomédico, el suceso es comúnmente denominado “muerte” y el tiempo como “tiempo de vida” Modelo de regresión de Cox y sus aplicaciones biosanitarias, Paola Velasco Álvarez, Universidad de Sevilla. (2015).

Para el presente trabajo, el tiempo de supervivencia se define como: el tiempo que un estudiante permanece en formación, medido a través de la condición de activo en el sistema de la IES, y el evento de interés es la deserción medida a partir del momento en que el estudiante presenta varios días con una condición de inactividad en el sistema.

El estimador de Kaplan-Meier (1958), de la función de supervivencia, es el estimador más común entre los paquetes estadísticos, este utiliza toda la información disponible, casos censurados y no censurados, para realizar la estimación de la función de supervivencia. El estimador en cualquier instante de tiempo es obtenido de la multiplicación de una secuencia de probabilidades condicionales de supervivencia estimadas. Cada probabilidad condicional estimada se obtiene del número de casos observados en riesgo y el número de “muertes” en un instante de tiempo y se calcula como $(n-d)/n$, donde n es el número de casos en riesgo y d es el número de “muertes” observadas. Análisis de Supervivencia Aplicado al Problema de la Deserción Estudiantil en la Universidad Tecnológica de Pereira. Mauricio Barrera Rebellon. Pereira (2008).

El análisis de supervivencia se basa en el procedimiento de Kaplan-Meier, quien se basa en la función Hazard y cuya distribución es Chi cuadrado es la prueba long Rank, en el caso que se desee comparar los dos modelos (H_0 : son iguales



El método de Kaplan-Meier (KM) es un método no paramétrico que se utiliza para estimar la probabilidad de supervivencia a partir de los tiempos de supervivencia observados (Kaplan y Meier, 1958).

El estimador de Kaplan y Meier que es utilizado para el caso en que se presente censura por la derecha es el siguiente:

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \left[\frac{r(t_i) - d(t_i)}{d(t_i)} \right]$$

Donde $d(t_i)$ representa el número total de ocurrencia en el momento t_i y $r(t_i)$ representa el número de unidades en riesgo, justo antes del tiempo t_i .

La probabilidad estimada $S(t)$ es una función escalonada que cambia de valor solo en el momento de cada evento. También es posible calcular intervalos de confianza para la probabilidad de supervivencia.

En el análisis de supervivencia el modelo de regresión más utilizado es el modelo de Cox, dada su flexibilidad y dado que, a la hora de interpretar los coeficientes, es algo más simple que el resto de los modelos propuestos. Este modelo, trabaja primordialmente con la función de riesgo de (Hazard function) y es utilizado para detectar relaciones existentes entre el riesgo que se produce en un determinado individuo en el estudio y algunas variables independientes y/o explicativas; por lo que este modelo nos permite evaluar dentro de un conjunto de variables cuales tienen relación, influencia...sobre la función de riesgo y por ello también en la función de supervivencia, ya que ambas funciones están conectadas. *Modelo de regresión de Cox y sus aplicaciones biosanitarias, Paola Velasco Álvarez, Universidad de Sevilla. (2015).*

El modelo de regresión de Cox a diferencia de otros modelos supone que existe un conjunto de covariables $X = (X_1, X_2, \dots, X_K)^T$, que influyen en el comportamiento del tiempo de ocurrencia de los eventos.

Este modelo trabaja bajo la función de riesgo $h[t|x_i(t)]$ que define el momento t dado un individuo con una característica en el tiempo t , $x_i(t)$. Dicha función de riesgo es de la forma:

$$h[t|x_i(t)] = h_0(t) [\exp(\beta'x_i(t))]$$



- 1) La expresión “ $\exp(\beta'x_i(t))$ ” se entiende como la parte paramétrica del modelo, y define el puntaje de riesgo (conocido como Risk Score), donde β es el vector de parámetros que miden el efecto de las covariables.
- 2) La expresión “ $h_0(t)$ ” se entiende como la parte no paramétrica del modelo, y define la función de riesgo base.

METODOLOGÍA

El alcance de esta investigación es de enfoque cuantitativo con una trascendencia de carácter descriptivo, teniendo en cuenta que ya que existen estudios e investigaciones sobre el ausentismo escolar, tanto a nivel local como regional, se realizaron técnicas de recolección de información cuantitativas de investigación, como encuesta sociodemográfica, identificación de un grupo focal y revisión documental; por otra parte, el enfoque también puede considerarse mixto, debido a que se realiza un estudio de caso no experimental, teniendo en cuenta, que se pretende explicar e identificar las causas más relevantes que conllevan a la deserción escolar de un grupo determinado de estudiantes. Villa Alzate Michael. (2017). Metodología en la Investigación. Disponible en <https://www.youtube.com/watch?v=8X2pdxYP7pk>.

La población del estudio está conformada por los estudiantes matriculados en la oferta académica 2019 comprendida entre el 26 de marzo de 2019 y 26 de marzo de 2020, de una IES de formación para el trabajo nivel técnico en la ciudad de Bogotá, pertenecientes a los estratos 1 y 2, el total de las mujeres corresponde a un 68% y los hombres al 32%, las edades se encuentran en un promedio entre los 18 y los 22 años.

La muestra es de carácter probabilística constituida por un total de 510 estudiantes, divididos en 17 grupos y en cuatro programas de formación, del total de los estudiantes desertaron 96 equivalente al 19%. La Coordinación Académica y el grupo de bienestar al estudiante facilitaron los datos, por tanto, se confirma la confiabilidad de la información

Fases de la Investigación:

1. Identificación del grupo focal

Para la identificación del grupo focal se verifico las ofertas academias anuales desde el 2015 hasta el 2019, identificando de esta manera que la deserción más alta en los últimos cuatro años fue la del 2019.

2. Aplicación de encuesta sociodemográfica.



La encuesta de sociodemográfica fue aplicada por la oficina de bienestar al aprendiz, a través de un formulario en línea, y la cual constaba de 26 preguntas cerradas.

3. Revisión exhaustiva documental.

- Se verificaron 96 actas de comité de evaluación y seguimiento ejecutados por la coordinación académica y corresponde a los 96 estudiantes que desertaron.
- Se verifico todos los datos disponibles en el proceso de ingreso, selección y matrícula de cada uno de los estudiantes.

4. Aplicación de instrumentos de análisis estadísticos

- Se realizó limpieza y depuración de la base de datos
- Se realizó análisis descriptivo exploratorio de las variables.
- Para estimar la función de sobrevivencia se implementó el estimador de Kaplan y Meier, de la función de supervivencia.
- Se elaboró el modelo semiparamétrico de regresión Cox, por ello también en la función de supervivencia, ya que ambas funciones están conectadas.

5. Análisis de resultados

El análisis de los resultados se realizó mediante los softwares: Excel, Word y Rstudio, la variable dependiente es la deserción; y las independientes son: Hijos, Motivo novedad, Programa, Alternativa, Estado de formación, Desnutrición, Edad, Peso, Estatura, Localidad, Estrato, Sexo, Condición social, Discapacidad, Estado civil, Dependientes, Apoyos estado, Trabajo actualmente, Nivel educativo mamá, Nivel educativo papá.

RESULTADOS

Análisis descriptivo exploratorio de los datos:

Tabla 1: Covariables Sexo, Estrato, Discapacidad.

ESTRAT O	FEMENINO			Total, Femenin o	MASCULINO				Total, Masculin o	Total, general			
	DISCAPACIDAD				DISCAPACIDAD								
	AUDITIV A	NINGUN A	VISUA L		DEL HABLA		MENTA L	NINGUN A					VISUA L
1	1	108	13	122	35%	0	0	37	12	49	30%	171	34%



2	2	199	16	217	62%	1	1	93	10	105	65%	322	63%
3	0	8	0	8	2%	0	0	8	0	8	5%	16	3%
4	0	1	0	1	0%	0	0	0	0		0%	1	0%
Total, general	3	316	29	348	100%	1	1	138	22	162	100%	510	100%

A partir de la Tabla 1 se obtiene que:

- ☐ Del total de estudiantes 348 son mujeres y 162 hombres
- ☐ El 34% pertenecen al estrato 1, el 63% al estrato 2
- ☐ 56 estudiantes tienen algún tipo de discapacidad correspondiente al 11%

Tabla 2: Covariables Estado Civil, Hijos

ESTADO_CIVIL	Número de Hijos					Total general	
	0	1	2	3	Más de 3		
CASADO/A	1	2	6	5	0	14	3%
SEPARADO/A	2	2	1	1	0	6	1%
SOLTERO/A	327	29	13	1	0	370	73%
UNION LIBRE	23	48	38	6	3	118	23%
VIUDO/A	0	0	0	2	0	2	0%
Total, general	353	81	58	15	3	510	1
	69%	16%	11%	3%	1%	100%	

De acuerdo con la tabla 2:

- ☐ El 73% de los estudiantes son solteros
- ☐ El 23% viven en unión libre
- ☐ El 3% son casados
- ☐ Solamente los estudiantes en unión libre tienen más de 3 de hijos
- ☒ Del total de los 157 hijos de los estudiantes, 95 son de los que viven en unión libre y 43 de los solteros, correspondiente al 61% y 27% respectivamente

Tabla 3: Covariables Programa de Formación, Alternativa de Etapa Productiva

PROGRAMA DE FORMACIÓN	CONTRATO DE APRENDIZAJE	NO SELECCIONO ALTERNATIVA	PASANTIA	PROYECTO PRODUCTIVO	VINCULO LABORAL	Total, general	
ASISTENCIA ADMINISTRATIVA	316	10	16	13	5	360	71%



CONTABILIZACION DE OPERACIONES COMERCIALES	50	1	1	6	2	60	12%
MANTENIMIENTO DE MOTORES GASOLINA Y GAS	14	0	8	8	0	30	6%
MANTENIMIENTO EQUIPO DE COMPUTO	50	1	5	4	0	60	12%
Total, general	430	12	30	31	7	510	100%
	84%	2%	6%	6%	1%	100%	

De acuerdo con la tabla 3:

- ☐ El 71% de los estudiantes pertenecen al programa Asistencia Administrativa
- ☐ El 84% de los estudiantes selección Contrato de Aprendizaje como Etapa Productiva
- ☐ El 2% no selecciono una alternativa de Etapa Productiva

Análisis de Supervivencia

Para iniciar el análisis se realiza mediante el estimador Kaplan y Meier para cada una de las variables, generando sus debidas curvas de supervivencia y el test de Log-Rank que permita identificar si la función de supervivencia para cada factor de las variables es igual o no.

Figura 1: Estimador de supervivencia covariable Trabajo actualmente.

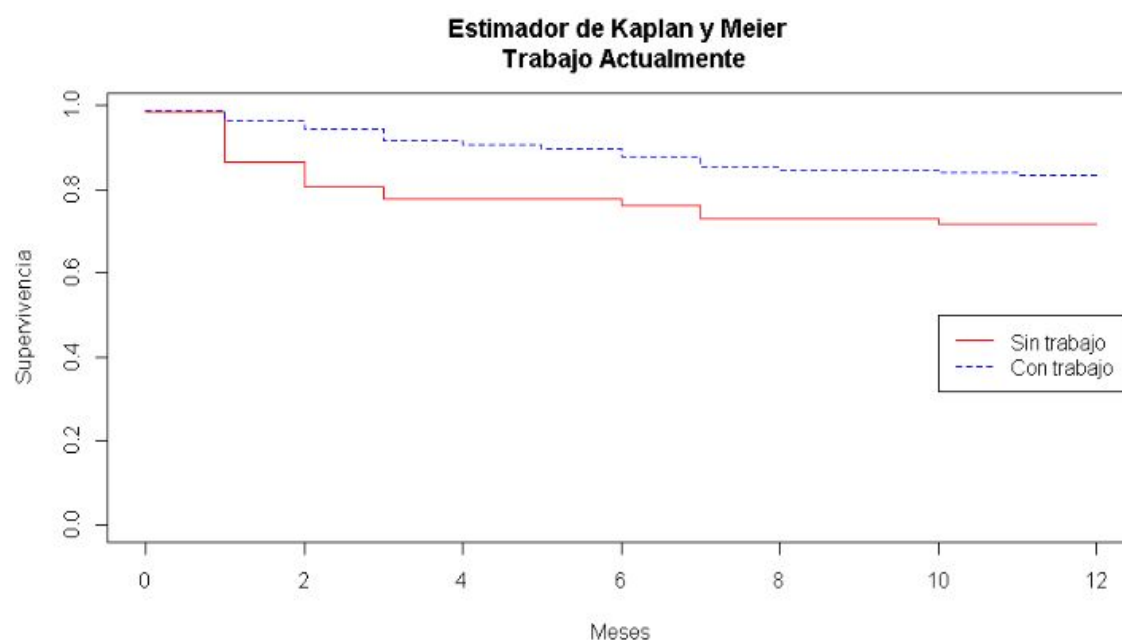


Tabla 4. Test Log-Rank para la covariable Trabajo Actualmente.



Trabajo Actualmente	Número de individuos	Eventos	Sobrevida	Límite inferior	Límite superior
Con trabajo	67	19	11.5	4.886	5.68
Sin Trabajo	443	77	84.5	0.665	5.68

Chisq= 5.7 on 1 degrees of freedom, $p = 0.02$

De acuerdo con la Figura 1, se puede evidenciar que existe diferencia en las curvas de supervivencia de los estudiantes con o sin trabajo; los estudiantes sin trabajo tienen una probabilidad aproximada del 79%, mientras que los que tienen trabajo del 82% de supervivencia, para confirmar esta teoría, y teniendo en cuenta el *p-valor* generado a través del test de Log-Rank en la Tabla No4, y bajo la hipótesis nula que asume que las dos funciones de supervivencia son iguales, hay evidencia para confirmar que estas curvas de supervivencia son diferentes con un nivel de significancia del 5%, lo que asegura que la covariable Trabajo Actualmente presenta un factor influyente al evaluar la deserción

Figura 2: Estimador de supervivencia covariable Alternativa de Etapa Productiva

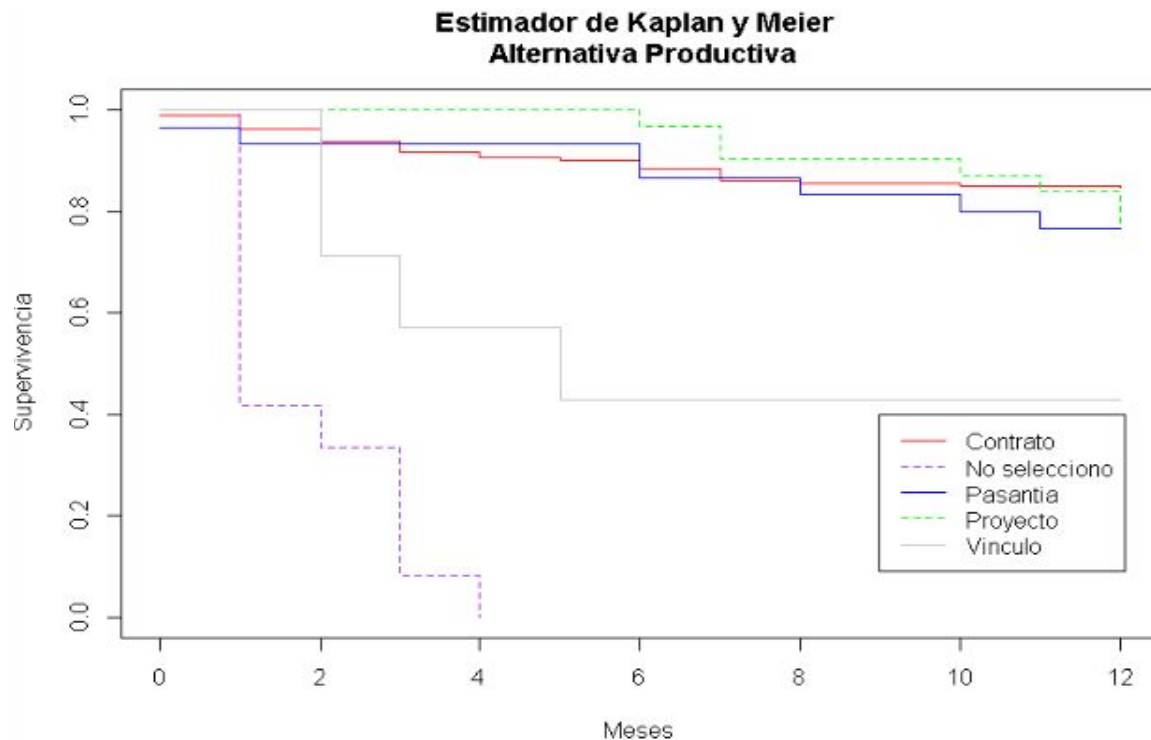




Tabla 5. Test Log-Rank para la variable Alternativa de Etapa Productiva.

Alternativa Etapa Productiva	Número de individuos	Eventos	Sobrevida	Límite inferior	Límite superior
Contrato de Aprendizaje	430	66	82.231	32.038	22.863
No selecciono Alternativa	12	12	0.857	1.448.579	152.003
Pasantía	30	7	5.665	0.3146	0.342
Proyecto Productivo	31	7	6.244	0.0914	0.100
Vínculo Laboral	7	4	1.002	89.652	9.287

Chisq= 164 on 4 degrees of freedom, $p < 2e-16$

De acuerdo con la Figura 2, se puede evidenciar que las curvas de supervivencia son distintas, de acuerdo con la Alternativa de Etapa Productiva Seleccionada, para confirmar está teoría y teniendo en cuenta el *p-valor* generado a través del test de Log-Rank en la Tabla 5, y bajo la hipótesis nula que asume que las dos funciones de supervivencia son iguales, hay evidencia para confirmar que estas curvas de supervivencia son diferentes con un nivel de significancia del 5%, lo que asegura que la variable Alternativa de Etapa Productiva presenta un factor influyente al evaluar la deserción.

Figura 3: Estimador de supervivencia covariable Estado Formación:



**Estimador de Kaplan y Meier
Estado Formación**

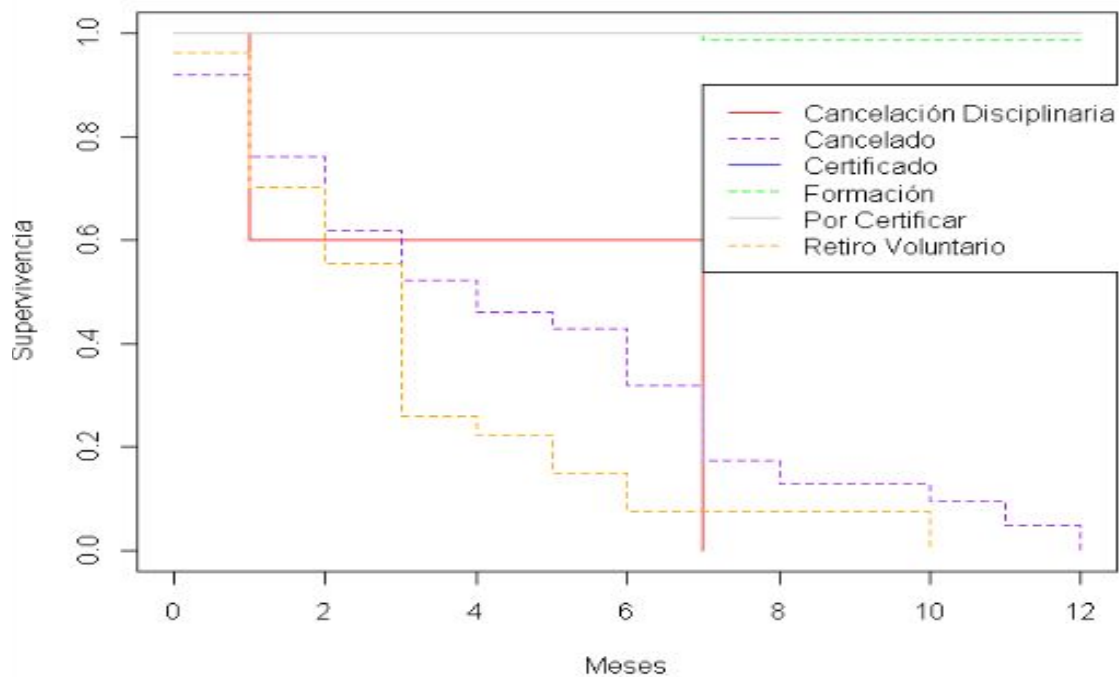


Tabla 6. Test Log-Rank para la variable Formación.

Estado Formación	Número de individuos	Eventos	Sobrevivida	Límite inferior	Límite superior
Cancelación Disciplinaria	5	5	0.625	30.6	31.7
Cancelado	63	63	7.337	422.3	476.6
Certificado	152	0	31.321	31.3	47.6
Formación	84	1	17.278	15.3	19.1
Por Certificar	179	0	36.885	36.9	61.4
Retiro Voluntario	27	27	2.554	234.0	250.6

Chisq= 817 on 5 degrees of freedom, $p < 2e-16$

De acuerdo con la Figura 3, se puede evidenciar que existe diferencia en las curvas de supervivencia de los estudiantes de acuerdo con el estado de la formación; para confirmar esta teoría, y teniendo en cuenta el *p-valor* generado a través del test de Log-Rank en la Tabla 6, y bajo la hipótesis nula que asume que las dos funciones de supervivencia son iguales, hay evidencia para confirmar que estas curvas de supervivencia son diferentes con



un nivel de significancia del 5%, lo que asegura que la covariable Estado Formación presenta un factor influyente al evaluar la deserción

Modelo de Regresión de COX

Tabla 7. Modelo de Regresión de Cox

Variables	Coefficiente de regresión	Exp (Coeficiente de regresión)	Error estándar (coeficiente de regresión)	z	Pr(> z)
Hijos	-0.1882563	0.8284024	0.2211487	-0.851	0.394622
Motivo novedad	-0.8923056	0.4097100	0.0684644	-13.033	< 2e-16 ***
Programa	-0.0892569	0.9146106	0.1434301	-0.622	0.533743
Alternativa	-0.3351303	0.7152449	0.1127729	-2.972	0.002961 **
Estado_ formación	18.128.448	61.278.555	0.1781874	10.174	< 2e-16 ***
Desnutrición	0.0064888	10.065.099	0.3209865	0.020	0.983872
Edad	0.0280882	10.284.864	0.0272962	1.029	0.303472
Peso	-0.0144367	0.9856670	0.0140050	-1.031	0.302623
Estatura	0.0181819	10.183.482	0.0219254	0.829	0.406957
Localidad	-0.2253411	0.7982439	0.0646068	-3.488	0.000487 ***
Estrato	-0.2912246	0.7473478	0.2743320	-1.062	0.288428
Sexo	0.2949666	13.430.814	0.4074505	0.724	0.469107
Codición_ social	0.0008821	10.008.825	0.3300163	0.003	0.997867
Discapacidad	0.2679989	13.073.457	0.3316113	0.808	0.418992
Estado civil	0.0328483	10.333.937	0.2569674	0.128	0.898283
Dependientes	0.1541971	11.667.208	0.1418262	1.087	0.276937
Apoyos estado	-0.0515003	0.9498034	0.1316611	-0.391	0.695681
Trabajo actualmente	-0.7009379	0.4961198	0.2929033	-2.393	0.016708 *
Nivel_ educativo_ mamá	0.0141519	10.142.525	0.0497938	0.284	0.776249
Nivel_ educativo_ papá	-0.0079701	0.9920616	0.0558814	-0.143	0.886586

Test de razón de verosimilitud	= 437.5 on 20 df, p=<2e-16
Test de Wald	= 209.8 on 20 df, p=<2e-16
Test de puntajes	= 736.7 on 20 df, p=<2e-16

De acuerdo con el Modelo de Cox realizado con anterioridad y los contrastes de hipótesis que permiten rechazar la hipótesis nula que representa la no significancia del modelo a un p-valor del 5%, se puede evidenciar que las siguientes variables representan un valor significativo al igual que el modelo.



- ☐ Motivo de La Novedad
- ☐ Alternativa
- ☐ Estado Formación
- ☐ Localidad
- ☐ Trabajo Actualmente

Teniendo en cuenta las anteriores covariables significativas, se procede a ajustar un nuevo modelo de Cox.

Tabla 8. Modelo de Regresión de Cox ajustado

Variable	Coefficiente de regresión	Exp (Coeficiente de regresión)	Error estándar (coeficiente de regresión)	z	Pr(> z)
Motivo novedad	-0.85208	0.42653	0.06113	-13.940	< 2e-16 ***
Alternativa	-0.31644	0.72874	0.10377	-3.049	0.00229 **
Estado_ formacion	175.608	578.969	0.15786	11.124	< 2e-16 ***
Localidad	-0.14914	0.86145	0.05111	-2.918	0.00353 **
Trabajo actualmente	-0.77371	0.46130	0.26493	-2.920	0.00350 **

Test de razón de verosimilitud = 428 on 5 df, p=<2e-16
--

Test de Wald = 218.2 on 5 df, p=<2e-16
--

test de puntajes = 727.9 on 5 df, p=<2e-16
--

De acuerdo con la tabla 8 donde se realiza el ajuste con las cinco variables significativas, y teniendo en cuenta los contrastes de las hipótesis con un p-valor del 5%, se puede concluir que el modelo general puede desarrollarse a partir de las covariables Motivo_novedad, Alternativa, Estado Formación, Localidad y Trabajo_actualmente.

Supuesto de riesgos proporcionales

El modelo de regresión de Cox también es conocido como el modelo de supuestos de riesgos proporcionales, dado que, las observaciones al ser independientes, el cociente entre el riesgo para dos individuos con el mismo vector de covariables es constante en el tiempo, y por lo tanto la proporcionalidad de los riesgos para los dos individuos no debe variar con respecto al tiempo. Por lo tanto, dicho supuesto se establece como:

$$\frac{h(x_i(t))}{h(x_j(t))} = \frac{h_0(t)[\exp(\beta'x_i(t))]}{h_0(t)[\exp(\beta'x_j(t))]} = \frac{\exp(\beta'x_i(t))}{\exp(\beta'x_j(t))}$$

El objetivo de dicho supuesto es el de no rechazar la hipótesis nula que representa el cumplimiento del supuesto de riesgo proporcional, lo que significa el vector de covariables es constante con respecto al tiempo. ("Análisis de Supervivencia Aplicado a la Deserción



en Estudiantes de la Facultad de Ingeniería de una Universidad de Bogotá, Colombia”. Sergio Yesid Aroca Pedroza. 2019).

Para determinar si las covariables anteriormente analizadas pueden permanecer constantes respecto al tiempo se ha puesto a prueba esto mediante el supuesto de riesgos proporcionales.

Tabla 9. Validación del supuesto de riesgos proporcionales

Covariable	chisq	Df	p
Motivo novedad	78.359	1	0.0051
Alternativa	0.0697	1	0.7918
Estado_formacion	37.985	1	0.0513
Localidad	53.736	1	0.0204
Trabajo_actualmente	0.8594	1	0.3539
GLOBAL	139.051	5	0.0162

Teniendo en cuenta la Tabla 9, y debido, a que a que el objetivo del Supuesto de Riesgos Proporcionales es el de no rechazar la hipótesis nula, se puede concluir, con un nivel de significancia del 5%, que el modelo cumple con el supuesto de riesgos proporcionales, a nivel global (exceptuando las covariables Motivo_novedad y Localidad). Esto significa que existe alguna tendencia en estas covariables con respecto al tiempo, y por lo tanto se puede afirmar que estas no son constantes. Esto último se puede corregir realizando una estratificación del modelo de Regresión de Cox.

Modelo de Cox estratificado

La estratificación de un modelo de regresión de Cox se efectúa, cuando no se cumple con el supuesto de riesgos proporcionales para una o más covariables, para esto se inicia estratificando aquella covariable que representa una significancia menor a las demás, para este caso la covariable Motivo_novedad y Localidad, donde al estratificarlas se contara a cada uno de los factores (Motivos de la novedad y Localidad) como estratos.

De manera inicial, se empieza realizando el modelo de regresión de Cox estratificados por las dos covariables que no cumplieron con el supuesto de riesgos proporcionales, esto con el fin de dar validez a este supuesto.

Estimación del segundo Modelo de Cox (Modelo definitivo) para la covariable Localidad

Tabla 10. Modelo de Regresión de Cox para las covariables estratificando por Localidad.



Covariable	Coefficiente de regresión	Exp (Coeficiente de regresión)	Error estándar (coeficiente de regresión)	z	Pr(> z)
Motivo novedad	-0.85247	0.42636	0.06682	-12.759	< 2e-16
Alternativa	-0.33542	0.71504	0.11062	-3.032	0.00243
Estado_formacion	172.045	558.706	0.17903	9.610	< 2e-16
Trabajo_actualmente	-0.53884	0.58343	0.29505	-1.826	0.06781

Test de razón de verosimilitud = 401.7 on 4 df, p=<2e-16
Test de Wald = 194.1 on 4 df, p=<2e-16
Test de puntajes = 688.7 on 4 df, p=<2e-16

De acuerdo con la Tabla 10 y después de estratificar el total de las variables por la covariable Localidad y teniendo en cuenta los contrastes de las hipótesis con un p-valor del 5%, se puede evidenciar que estas son significativas.

Estimación del segundo Modelo de Cox (Modelo definitivo) para la covariable Motivo_novedad

Tabla 11. Modelo de Regresión de Cox para las covariables estratificando por Motivo_novedad.

Covariable	Coefficiente de regresión	Exp (Coeficiente de regresión)	Error estándar (coeficiente de regresión)	z	Pr(> z)
Alternativa	-0.09377	0.91049	0.10120	-0.927	0.354
Estado_formacion	-0.06513	0.93695	0.33438	-0.195	0.846
Trabajo_actualmente	-0.32911	0.71957	0.28574	-1.152	0.249
Localidad	-0.08256	0.92075	0.05789	-1.426	0.154

Test de razón de verosimilitud = 4.44 on 4 df, p=0.3
Test de Wald = 4.26 on 4 df, p=0.4
test de puntajes = 4.33 on 4 df, p=0.4

De acuerdo con la Tabla 11, y después de estratificar el total de las variables por la covariable Motivo_novedad y teniendo en cuenta los contrastes de las hipótesis con un p-valor del 5%, se puede evidenciar que estas no son significativas, ni por el modelo en general ni por covariables individuales.

Por tanto, se puede identificar que, al estratificar la covariable Localidad, y teniendo en cuenta los contrastes de las hipótesis con un p-valor del 5%, que busca rechazar la hipótesis nula que representa la no significancia del modelo, se confirma que este al igual que sus covariables de forma independiente son significativas.



Ahora bien, para determinar si las covariables anteriores del modelo definitivo de Cox pueden permanecer constantes respecto al tiempo, se proceder a realizar nuevamente la prueba del supuesto de riesgos proporcionales.

Supuesto de riesgos proporcionales

Tabla 12. Supuestos de riesgos proporcionales después de estratificar por la covariable Localidad

Covariables	chisq	df	p
Motivo_novedad	62.033	1	0.013
Alternativa	0.9638	1	0.326
Estado_formacion	30.643	1	0.080
Trabajo_actualmente	0.0355	1	0.851
GLOBAL	80.345	4	0.090

Considerando los *p-valores* generados en la Tabla 12 y debido a que el objetivo del Supuesto de Riesgos Proporcionales es el de no rechazar la hipótesis nula, se puede concluir, con un nivel de significancia del 5% que el modelo cumple con el supuesto de riesgos proporcionales, a nivel global, por covariable encontramos que el Motivo_novedad no es significativa debido a que tiene demasiadas posibilidades que hacen que esta no se comporte constante en el tiempo.

Análisis de residuales

Una de las ventajas que han surgido del enfoque del análisis de supervivencia es la posibilidad de efectuar análisis de residuos (Andersen et al., 1993, Fleming y Harrington, 1991, Therneau y Grambsch, 2000, Therneau et al., 1990). Borges, R. (2004). *Análisis de Supervivencia básico utilizando el lenguaje R*

Los residuos se pueden utilizar para:

1. Descubrir la forma funcional correcta de un predictor continuo.
2. Identificar los sujetos que están pobremente predichos por el modelo.
3. Identificar los puntos o individuos de influencia.
4. Verificar el supuesto de riesgo proporcional.

Los residuos de Schoenfeld son útiles para la verificación del supuesto de riesgo proporcional en el modelo de Cox.

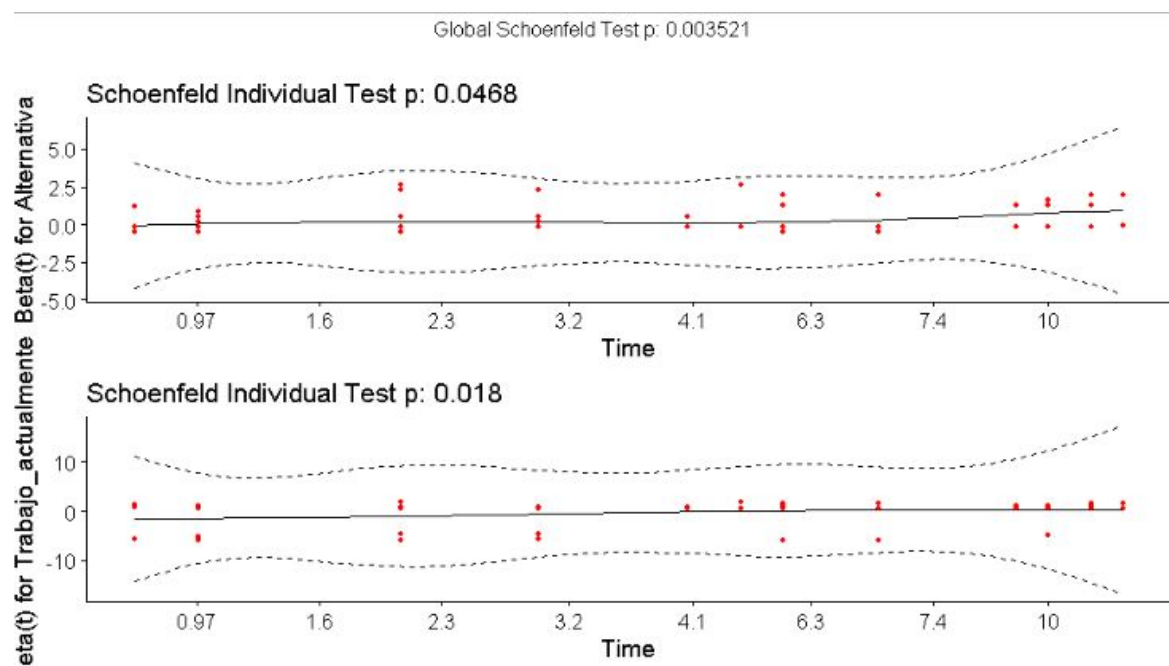


Los residuos de Schoenfeld (1982) se define como la matriz:

$$s_{ij}(\beta) = Z_{ij}(t_i) - Z_j(\beta, t_i)$$

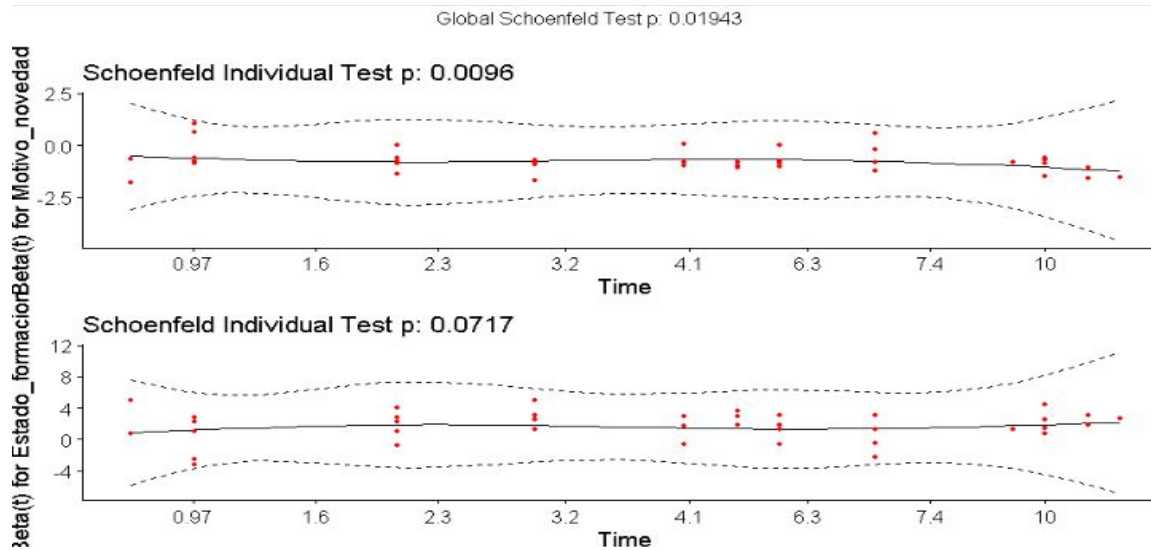
con una fila por muerte y una columna por covariable, donde i y j son los individuos y el tiempo de ocurrencia del evento respectivamente.

Figura 4. análisis de residuales de Schoenfeld del supuesto de riesgos proporcionales del modelo de regresión de Cox propuesto en la tabla 10, covariables Trabajo actualmente y Alternativa de Etapa Productiva



De acuerdo con la figura 4, se puede evidenciar que las variables Motivo_novedad y Estado formación, se observa un tipo de tendencia en la curva ajustada (línea continua) entre tiempo y residuos, lo que da indicios de la verificación de la hipótesis de riesgos proporcionales.

Figura 5. análisis de residuales de Schoenfeld del supuesto de riesgos proporcionales del modelo de regresión de Cox propuesto en la tabla 10, covariables Estado Formación y Motivo Novedad.



De acuerdo con la figura 5, se evidencian que respecto a las covariables Alternativa y Trabajo actualmente, al agruparse los residuos de forma aleatoria a ambos lados del valor 0 del eje y, y al no observarse una tendencia con cambios bruscos, entonces puede observarse que no hay violación del supuesto de riesgo proporcional

CONCLUSIONES:

Realizado el análisis de supervivencia por medio de los estimadores de Kaplan y Meier, y el modelo de Riesgos Proporcionales de Cox, aplicado y ajustado a la población total de 510 estudiantes de una IES de nivel Técnico en la ciudad de Bogotá, se obtuvieron los siguientes resultados:

Las covariables que probablemente tendrían cierto peso en el estudio, como los son: tener hijos, personas dependientes, algún grado de desnutrición, cierta edad, peso, estatura, alguna discapacidad, la condición social, el estado civil o simplemente el hecho de ser hombre o mujer, no tienen significancia en el modelo, lo que quiere decir que los estudiantes de la oferta académica 2019 de la IES en estudio, en su mayoría no abandonaron la formación por alguna de estas razones.

De acuerdo con la covariable alternativa de etapa productiva, tienen más probabilidad de desertar en los primeros cuatro meses de formación los estudiantes que no seleccionaron ningún tipo de alternativa, los estudiantes que tomaron el vínculo laboral abandonaron hasta el quinto mes de formación. Iniciada la etapa productiva en el sexto mes ningún aprendiz deserta.

En la covariable estado de la formación, se evidenció que los estudiantes que se retiraron de forma voluntaria lo realizaron hasta el sexto mes de la formación, de ahí en adelante se mantuvo constante sin retiros hasta el décimo mes de la formación; el factor “cancelado” se mantuvo constante desde el inicio hasta el final de la formación.



De acuerdo al modelo de Riesgos Proporcionales, al análisis de supuesto de Riesgos y posteriormente a la estratificación de la variable localidad, se pudo establecer que las covariables Alternativa, Estado_ formación y Trabajo_ actualmente explican el modelo; por otra parte, la covariable Motivo_ novedad, aunque se muestra significativa en el modelo de Cox, no se mantiene en el análisis de Riesgos Proporcionales, debido a esto se recomienda realizar un estudio más profundo teniendo esta variable como principal motivo de análisis.

El hecho de haber estratificado la covariable Localidad justifica realizar el supuesto de riesgos proporcionales, donde se ha demostrado que Localidad es dependientes del tiempo; una vez realizada la estratificación se demostró que las covariables cumplían con el supuesto, lo que permite determinar que ahora son constantes con respecto al tiempo. Esto demuestra que el análisis es más funcional si se toma a cada localidad como un estrato diferente entre estos.

REFERENCIAS BIBLIOGRÁFICAS

Arteaga González, A., Torres Ávila, K., & López Cardona, L. D. (2016). *Variables asociadas al riesgo de deserción y su relación con los beneficios y servicios ofrecidos por Bienestar al Aprendiz en el Centro de Servicios Financieros. Revista Finnova.2(4), 65 - 86.*

Cely Atuesta, D., & Durán Gamba, M. (2014). *Causas Asociadas A La Deserción Estudiantil Y Estrategias De Acompañamiento Para La Permanencia Estudiantil. Congresos CLABES.*

Gardner Isaza, Lorena; Dussán Lubert, Carmen; Montoya Londoño, Diana Marcela. (2012-2014). *Aproximación causal al estudio de la deserción en la Universidad de Caldas. Periodo*

Servicio Nacional de Aprendizaje, (2020). *Guía para la elaboración del plan de acción de bienestar al aprendiz en el Centro de Formación.*

Cárdenas Leuro, Mariana (2013) *Un modelo de sobrevida multivariado para eventos recurrentes por sujeto con evento terminal: deserción de clientes en la industria de las telecomunicaciones.*

Ocampo Mesa Leidy Alejandra. (2019) *Modelo de sobrevida aplicado a deserción estudiantil de las facultades de Estadística y Diseño gráfico de la Universidad Santo Tomás.*

Velasco Álvarez Paola. (2015/16 Universidad de Sevilla) *Modelo de regresión de Cox y sus aplicaciones biosanitarias*



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

Raffino María Estela. Disponible en: <https://concepto.de/desercion-escolar/>. Consultado: 29 de agosto de 2020). *Deserción Escolar*.

Universidad de los Andes, 2014, citado por (Gartner Isaza, Lorena y Dussán Lubert, Carmen y Montoya, Diana Marcela (2016). *Caracterización de la Deserción Estudiantil en la Universidad de Caldas el Período 2009-2013*

Aroca Pedroza Sergio Yesid. (2020) *Análisis de Supervivencia Aplicado a la Deserción en Estudiantes de la Facultad de Ingeniería de una Universidad de Bogotá, Colombia*

Barrera, M. (2008). *Análisis de supervivencia aplicado al problema de la deserción estudiantil en la Universidad Tecnológica de Pereira*. Pereira, Colombia: Universidad Tecnológica de Pereira.

Borges, R. (2004). *Análisis de Supervivencia básico utilizando el lenguaje R*. Caracas, Venezuela: Universidad Central de Venezuela. Escuela de Estadística y Ciencias Actuariales